THE TIP OF THE ICEBERG

Say, what is "under the floor" of a supercomputer?

Every year for the past 17 years, the director of Los Alamos National Laboratory has had a legally required task: write a letter—a personal assessment of Los Alamos–designed warheads and bombs in the U.S. nuclear stockpile. This letter is sent to the secretaries of Energy and Defense and to the Nuclear Weapons Council. Through them the letter goes to the president of the United States.

The technical basis for the director's assessment comes from the Laboratory's ongoing execution of the nation's Stockpile Stewardship Program; Los Alamos' mission is to study its portion of the aging stockpile, find any problems, and address them. And for the past 17 years, the director's letter has said, in effect, that any problems that have arisen in Los Alamos weapons are being addressed and resolved without the need for full-scale underground nuclear testing.

When it comes to the Laboratory's work on the annual assessment, the director's letter is just the tip of the iceberg. The director composes the letter with the expert advice of the Laboratory's nuclear weapons experts, who, in turn, depend on the results from another year's worth of intense scientific investigation and analysis done across the 36 square miles of Laboratory property.

One key component of all that work, the one that the director and the Laboratory's experts depend on to an ever-increasing degree, is the Laboratory's supercomputers. In the absence of real-world testing, supercomputers provide the only viable alternative for assessing the safety, reliability, and performance of the stockpile: virtual-world simulations.

I, Iceberg

Hollywood movies such as the *Matrix* series or *I, Robot* typically portray supercomputers as massive, room-filling machines that churn out answers to the most complex questions—all by themselves. In fact, like the director's Annual Assessment Letter, supercomputers are themselves the tip of an iceberg.

Without people, a supercomputer would be no more than a humble jumble of wires, bits, and boxes.

Although these rows of huge machines are the most visible component of supercomputing, they are but one leg of today's supercomputing environment, which has three main components. The first leg is the supercomputers, which are the processors that run the simulations. The triad also includes a huge, separate system for storing simulation data (and other data). This leg is composed of racks of shelves containing thousands of data-storage disks sealed inside temperature- and humidity-controlled automated libraries. Remotely controlled robotic "librarians" are sent to retrieve the desired disks or return them to the shelves after they are

The most important assets in the Laboratory's supercomputing environment are the people—designing, building, programming, and maintaining the computers that have become such a critical part of national security science. (Photo: Los Alamos)

played on the libraries' disk readers. The third leg consists of the many *non*-supercomputers at the national security laboratories. The users of these computers request their data, over specially designed networks, from the robotic librarians so they can visualize and analyze the simulations from afar.

The Los Alamos supercomputers are supported by a grand infrastructure of equipment used to cool the supercomputers and to feed them the enormous amounts of electrical power they need. They also need vast amounts of experimental data as input for the simulation codes they run, along with the simulation codes themselves (also called programs, or applications), tailored to run efficiently on the supercomputers. In addition, system software is necessary to execute the codes, manage the flow of work, and store and analyze data.

People are the most vital component of any supercomputer's supporting infrastructure. It takes hundreds of computer scientists, engineers, and support staff to design, build, maintain, and operate a supercomputer and all the system software and codes it takes to do valuable science. Without such people, a supercomputer would be no more than a humble jumble of wires, bits, and boxes.

The computer room's vast floor space is 43,500 square feet, essentially an acre—90 percent of a football field.

Supercomputers That Fill a Stadium

At Los Alamos, supercomputers, and the immense amount of machinery that backs them up, are in the Nicholas C. Metropolis Center for Modeling and Simulation, known pragmatically as the Strategic Computing Center (SCC).

Roadrunner, the world's first petaflop computer, joined other supercomputers in the SCC's computer room in 2008. It is a *big* machine, containing 57 miles of fiber-optic cables and weighing a half-million pounds. It covers over 6,000 square feet of floor space, 1,200 square feet more than a football field's end zone. But that represents only a portion of the computer room's vast floor space, which is 43,500 square feet, essentially an acre—90 percent of a football field (minus the end zones). (Roadrunner has finished its work for the Laboratory and is currently being shut down.)

What is really amazing, however, lies *beneath* the supercomputer room floor. A trip straight down reveals more vast spaces crowded with machinery that users never see.

The computer room is the SCC's second floor, but that one floor is actually *two*, separated by almost four feet. That 4-foot space hosts the miles of bundled network cables, electrical power lines inside large-diameter conduit, and other subfloor equipment the supercomputers rely on. The double floor provides enough room for engineers and

The Metropolis Center, also called the Strategic Computing Center, is the home of Los Alamos' supercomputers and the vast infrastructure that supports them. (Photo: Los Alamos)

maintenance staff, decked out like spelunkers in hardhats and headlamps, to build and manage these subfloor systems.

Below this double floor, on the building's first floor, is another acre-size room, a half-acre of which holds row upon row of cabin-size air-conditioning units. These cool the air and then blow it upwards into the computing room, where it draws the heat off the hard-working computers. The now-warmed air then rises to the third floor (basically an acre of empty space), whereupon it is drawn back down, at the rate of 2.5 million cubic feet per minute, to the first floor by the air coolers so the cooling cycle can begin again.

An additional half-acre of floor space stretches beyond the cooling room and holds the SCC's electric power infrastructure, the machines that collectively keep the supercomputers running. There are rows of towering power distribution units (PDUs), containing transformers and circuit breakers, and for backup power, rotary uninterruptible power supply (RUPS) generators. Each RUPS uses motor generator technology. Electricity fed into the RUPS is used to build kinetic energy in a 9-foot-diameter flywheel that, in turn, generates electricity.

Supercomputers are cooled by chilled air circulating at the rate of 2.5 million cubic feet per minute.

That bit of extra electricity evens out the flow of power to the supercomputers in the case of a power surge from, for example, a lightning strike, a common occurrence in summertime Los Alamos. In the case of a power outage, there is enough kinetic energy built up in the flywheel to provide 8–12 seconds of electricity to the supercomputers. Those few seconds are long enough for data about the current state of a running calculation to be written to memory, reducing the loss of valuable data.

The PDUs transform the incoming high-voltage electric power feed into lower voltage and distribute it to the supercomputers according to each machine's particular voltage needs, for example, 220 volts for Roadrunner and 480 volts for the supercomputer named Cielo.

The Guardians

Because the Laboratory's supercomputers work on national security problems 24 hours a day, 365 days a year, they require dedicated overseers who stay onsite and collectively keep the same exhausting schedule. The members of the SCC's operations staff, a team of 22, are the experts who keep things running and make sure anything that goes wrong gets fixed, right away.

Divided into three shifts, the members of the operations staff tend monitoring equipment and keep watch from inside the Operations Center, a high, windowed nerve center that overlooks the computer room. The staff's tasks are many and varied, as they are charged not only with overseeing the computer hardware and software but also, for example, with keeping tabs on the cooling system. The computer room's environment must stay cool enough to prevent damage to the valuable computers; too much heat is a major threat.

These dedicated guardians are expected to be able to fix both hardware and software problems in about an hour. For software problems requiring additional support, a team of 30 software administrators, also stationed onsite, backs them up. If a software problem occurs outside regular business hours, the administrators can be called in and must report to the SCC within two hours.

Evolution to Revolution

Looking for all the world like row upon row of large gym lockers, a supercomputer is visibly very different from a personal computer (PC). But the real difference is in the work supercomputers do and the way they do it.

The guardians are expected to be able to fix both hardware and software problems in about an hour.

Today's supercomputers are collections of tens of thousands of processors housed in "racks," cabinets holding the processors and supporting equipment. The large number of processors is needed because supercomputers run immense calculations that no PC could do. The calculations are divided into smaller portions that the processors work on concurrently. This is parallel computing or actually, for a supercomputer, *massively* parallel computing.

A new supercomputer for Los Alamos can take years to create. The process begins with an intense collaboration between commercial computer companies, like IBM,

The Laboratory's Luna supercomputer can be accessed by all three national security labs (Los Alamos, Livermore, and Sandia), making it a "trilab" machine. Roadrunner and Cielo are also trilab machines. (Photo: Los Alamos)

Cray, Hewlett-Packard, etc., and Los Alamos' computer experts, who have extensive experience both operating and designing supercomputers. Los Alamos computer personnel involve themselves in the creation of each new Laboratory supercomputer from the generation of the first ideas to the machine's delivery . . . and after its delivery. Once it is built and delivered, before it is put to work, a supercomputer is disassembled, inspected, and reassembled to ensure that it can handle classified data securely and can be fixed and maintained by Laboratory staff.

Once it is built and delivered, a supercomputer is disassembled, inspected, and reassembled to ensure that it can handle classified data securely.

As a practical and economic necessity, each new Los Alamos supercomputer takes advantage of commercial technological advances. And in the 21st century, beginning with Roadrunner, technology from the private sector is being evolved in innovative ways that are, in effect, a reinvention of how a supercomputer is built. Roadrunner, for example, used video game technology originally conceived for the Sony PlayStation 3, and with that technology, it became the world's first hybrid supercomputer, with an architecture that linked two different types of processors to share computational functions. This particular evolutionary step in supercomputer architecture let Roadrunner surge onto the global stage as the world's first petaflop computer.

Architectures are still evolving, so the next generation of machines will be radically new, even revolutionary, as will Trinity, Los Alamos' next supercomputer, projected to arrive in 2015–2016. On Trinity, Laboratory designers and their industry partners will be trying out numerous innovations that will directly affect supercomputing's future. So Trinity will be unlike any other computer Los Alamos researchers have used. And by the way, it will be 40 to 50 times faster than Roadrunner.

The exact form Trinity will take is still being decided, as design discussions are still underway, but whatever the final design is, it will be a means to an end. The form each new supercomputer takes is dictated by what the Laboratory needs the machine to do. In general that always means it must answer more questions, answer new kinds of questions about new and bigger problems, compute more data, and compute more data faster.

Los Alamos' specific need, however, is focused on the stockpiled nuclear weapons and the continuous analysis of them. Laboratory supercomputers are already simulating the detonation of nuclear weapons, but Trinity and the computers that will succeed it at the Laboratory will need to simulate more and more of the entire weapon (button-to-boom) and in the finest-possible detail. Design efforts for Trinity will be aimed at that goal, and a great deal of effort will go into creating the many new and complex subsystems that the computer will need.

Saving Checkpoints Is the Name of the Game

At the system level, some design requirements remain the same from supercomputer to supercomputer, even when the next one is as fundamentally different as Trinity will be. For example, while a PC serves one user at a time, Laboratory supercomputers must serve many users simultaneously—users from the Laboratory's various divisions and from the other national security labs far beyond Los Alamos. The computer they use must be designed not only to accommodate that multitude of users but also to provide ultra-secure access for the protection of classified data.

Every Los Alamos supercomputer must also be designed to enable an operator to quickly and easily identify and locate which component within the computer's 6,000 square feet (or more) of equipment needs repair. And repairs will always be needed because of the ever-increasing size and speed of supercomputers. As these machines get larger and faster, they naturally become more and more subject to breakdown.

Think about this: If a PC crashes once year and a supercomputer is equal to at least 10,000 PCs, one might expect to see 11 failures *per hour* on a supercomputer. Consider what such a failure rate could mean for an extensive computation. At Los Alamos, a nuclear weapon simulation can take weeks or even months to be completed, and those weeks and months are already costly in terms of computer time filled and electrical power used. In addition, successful simulations require a large collaborative effort between, for example, the weapons scientists, computer designers, computer code developers, and members of the supercomputer operations team. A breakdown equals time and money lost.

With downtime being a supercomputing inevitability, it is commonplace to mitigate the loss by "checkpointing," which is like hitting "Save." At predetermined times—say, every four hours—the calculation is paused and the results of the computation up to that point (the "checkpoint") are downloaded to memory. Returning the simulation to the closest checkpoint allows a simulation (or other type of calculation) to be restarted after a crash with the least amount of data loss.

Unfortunately, the compute time lost even to checkpointing is becoming dearer as supercomputers grow larger and therefore more prone to periodic crashes, so Trinity's designers are working on new checkpointing methods and systems that will maintain a higher level of computational productivity. Los Alamos is working closely with industry to develop this kind of defensive capability.

An Itch That Needs Scratching

PCs are all fundamentally the same, similarly designed to do the same tasks. Users can just go out and buy the software they need for their brand of PC. But supercomputers are different. Designed and built to fill a specific need, each one scratches a hard-to-reach itch. At Los Alamos, the special need is scientific computing and simulation, and a super-computer's users need specially written codes for each project.

Who develops the advanced codes used on Los Alamos supercomputers—the codes for weapon simulation or for general science research? Those highly specialized programs are created in-house, and for many years, the Laboratory's successive supercomputers have had enough in common that existing codes adapted well to them. Trinity's architecture and performance characteristics, however, will presage a complete upheaval. The codes will need to be overhauled, not just adapted: more of a "build it from scratch" compared with just an updating.

The developers are already busy making codes "Trinity friendly" and doing so without having anywhere near the variety and amount of resources the giant commercial computer companies have available. For this work, developers depend on partnering with a range of Laboratory scientists, who provide the unique algorithms for solving the basic physics equations governing how the dynamics of a complex system play out over time. This is true whether the system being studied is the climate or a nuclear explosion. The nature of the scientists' algorithms and the new data generated as a system changes with time determine how the code developers design and build a code to make efficient use of the supercomputer and its data storage and networking connections. In this age of "big data," building programs that efficiently generate unbelievably massive datasets on a supercomputer and make them useful has become a grand challenge. (See the article "Big Data, Fast Data—Prepping for Exascale" in this issue.)

A Titanic Achievement

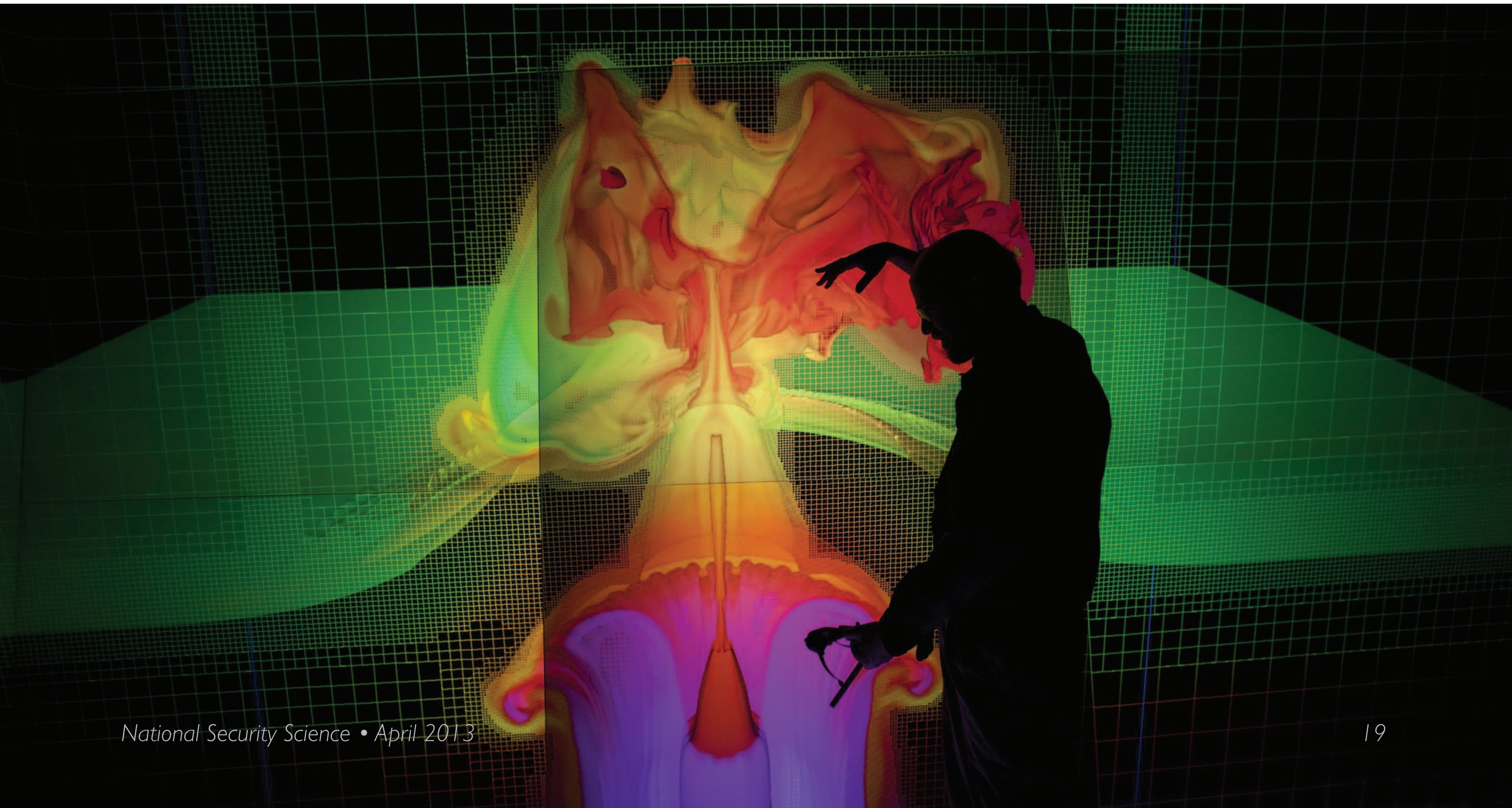
Designing, building, operating, and maintaining a supercomputer are completely different experiences than working with Word or Excel on a PC at home or at the office. That is true today and will be true, in spades, tomorrow. Computer architectures continue to evolve, leading to the upcoming Trinity and eventually to machines still unimagined.

The Laboratory's supercomputers cannot exist without massively complex and expensive infrastructures, which are often unacknowledged and unappreciated, and without the effort and creative thinking of hundreds of scientists, engineers, and technicians. Working together, they meet the challenge of providing the most-advanced supercomputing environments in the world and then use them to perform the national security science that makes the director's Annual Assessment Letter possible.

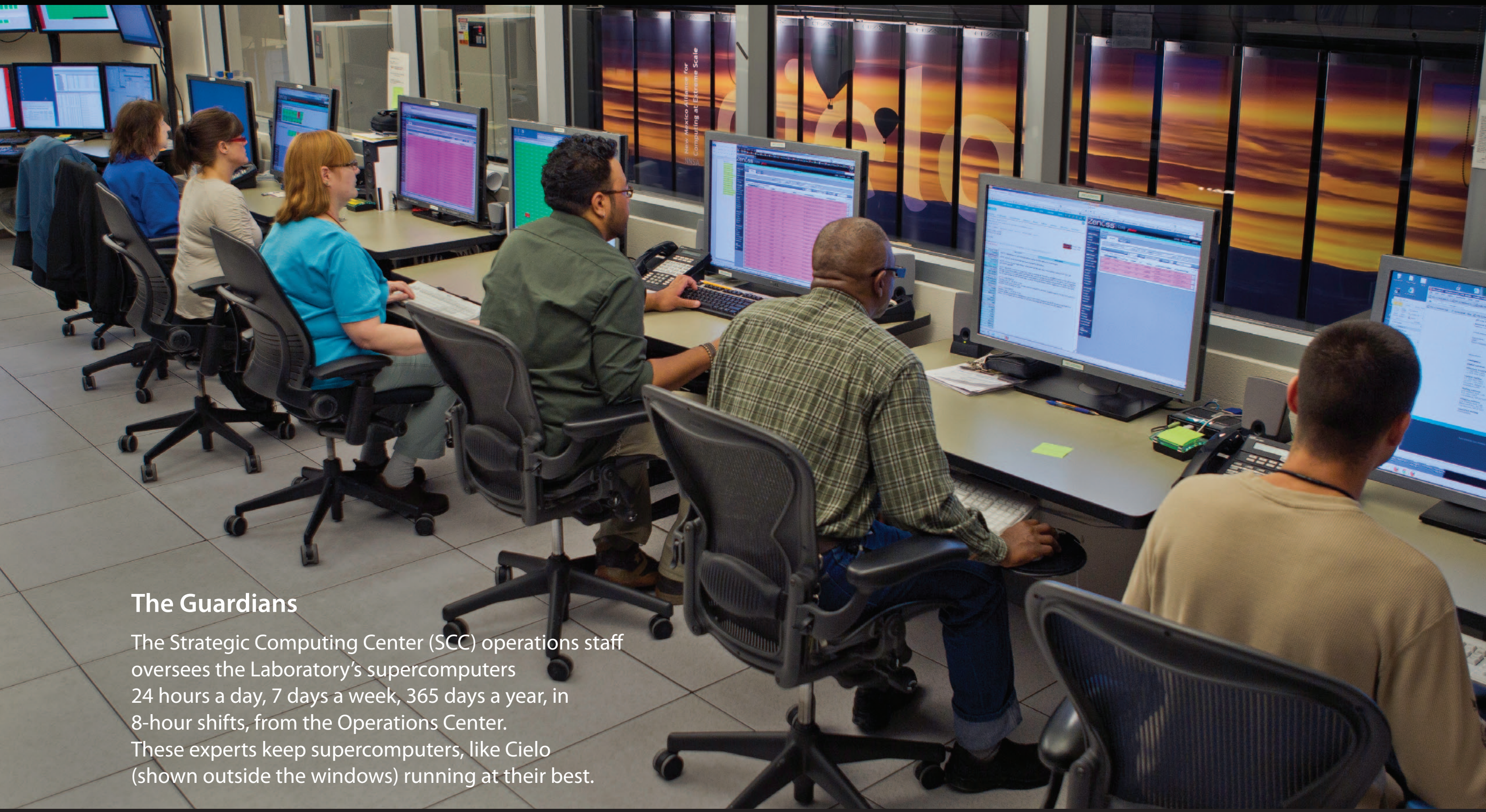
It is hard work, and it is certainly worth it.

~ Clay Dillingham

Using supercomputers, scientists can interact with simulations of everything from nuclear detonations to protein synthesis or the birth of galaxies. These simulations can boggle the mind—and at the same time provide clarity. Scientists wear special glasses to view simulations in 3D at extremely high resolution. They can even manipulate the simulations, as the viewer shown here is doing. (Photo: Los Alamos)



What it takes to run a roadrunner

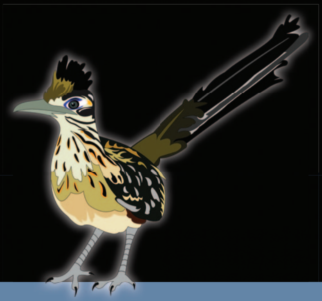


The Guardians

The Strategic Computing Center (SCC) operations staff oversees the Laboratory's supercomputers 24 hours a day, 7 days a week, 365 days a year, in 8-hour shifts, from the Operations Center. These experts keep supercomputers, like Cielo (shown outside the windows) running at their best.

Floor Space

The SCC is a 300,000-square-foot building. The vast floor of the supercomputing room is 43,500 square feet, almost an acre in size.



Electric Power

The amount and cost of electric power required to run a supercomputer are staggering.

Today, a megawatt (MW) of power costs $1 million per year. Roadrunner uses 2 MW per year. Cielo, the Laboratory's newest supercomputer, is a 3-MW machine. Trinity will be a 12-MW machine.

The combined supercomputing facilities at Los Alamos use $17 million per year of electricity.

Using all that electric power means that supercomputers generate lots of heat. If not kept cool, a supercomputer will get too hot and overheat, causing processors to fail and the machine to need costly, timely repairs.

An electrician, wearing personal protective gear, works on a 480-volt breaker inside a power distribution unit.

Managers at the SCC inspect the double floor beneath the acre-size supercomputing room. Several of the giant air-cooling units are visible in the foreground and behind the managers.

The high winds blowing beneath the supercomputer room are generated by the massive air-cooling units.

To capture the dust and dirt that might otherwise blow into the supercomputers, the 84 giant air-coolers use 1,848 air filters. It takes two staff members an entire month to change the filters.

Air-Cooling

Beneath the acre-size supercomputing room in the SCC is a 1.5-acre floor that houses 84 giant 40-ton air-cooling units. Together, these units can move 2.5 million cubic feet of chilled air per minute through the supercomputing room above.

The air-cooling units use water, cooled by evaporation, to chill the air before it is blown upward to circulate around the supercomputers.

The air, now heated by cooling the supercomputers, is drawn back down to the lower floor and back into the air-cooling units. This process transfers the heat from the air to the water, which is then recooled by evaporation.

Water for Cooling

The amount of water required to cool the air that, in turn, cools a supercomputer is also staggering. The SCC uses 45,200,000 gallons of water per year to cool its supercomputers. This amount of water costs approximately $120,000 per year.

By the end of the decade, as supercomputers become more powerful and require more cooling, the SCC is predicted to double its water use to 100,000,000 gallons.

The SCC has five evaporative cooling towers. These towers evaporate water to dissipate the heat absorbed by the water in the air-cooling units.

There is room to add an additional cooling tower as the supercomputing needs of the Stockpile Stewardship Program increase.